U N I V E R S I T Ä T   H A M B U R G

A note on residual-based empirical likelihood kernel density estimation

Birte Muhsal and Natalie Neumeyer

Preprint–No. 2010-05   May 2010

DEPARTMENT MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

A note on residual-based empirical likelihood kernel density estimation

Birte Muhsal
Department of Mathematics
Karlsruhe Institute of Technology

Natalie Neumeyer *
Department of Mathematics
University of Hamburg

May 31, 2010

Abstract

In general the empirical likelihood method can improve the performance of estimators by including additional information about the underlying data distribution. Application of the method to kernel density estimation based on independent and identically distributed data always improves the estimation in second order. In this paper we consider estimation of the error density in nonparametric regression by residual-based kernel estimation. We investigate whether the estimator is improved when additional information is included by the empirical likelihood method. In comparison to the residual-based kernel estimator we observe a change in the asymptotic bias of the empirical likelihood estimator already in first order and in the asymptotic variance in second order.

AMS 2010 Classification: 62G07, 62G08

Keywords and Phrases: asymptotic mean squared error, error distribution, kernel estimation, likelihood method, nonparametric regression, second order expansions

1 Introduction

Let $\varepsilon_1, \ldots, \varepsilon_n$ denote a sample of independent random variables with cumulative distribution function F and density f. In nonparametric inference the distribution function F is typically estimated

*corresponding author, e-mail: neumeyer@math.uni-hamburg.de, address: Fachbereich Mathematik, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany, Tel. +49 (0)40 42838 4907, Fax. +49 (0)40 42838 6955

by the empirical distribution function F_n, and the density f by a kernel estimator f_n [see Rosenblatt (1956), among many others]. Now assume that additional information about the distribution of interest is available in form of the equality

$$E[g(\varepsilon_1)] = \int g(y)f(y)\,dy = 0, \tag{1.1}$$

for some known function g, e.g. $g(y) = (y, y^2 - \sigma^2)^\top$ for centeredness and a known variance σ^2. This information can be incorporated into the estimation by the empirical likelihood method introduced by Owen (2001) [see also Hall and LaScala (1990), DiCiccio et al. (1989), Kitamura (1997), Einmahl and McKeague (2003), and Hjort et al. (2009), among many others]. It is shown by Qin and Lawless (1994) that the empirical likelihood estimator for F, say $\tilde{F}_n$, in comparison to F_n always has a smaller asymptotic mean squared error. Whereas for the estimation of the distribution function the improvement by the empirical likelihood method is observable in the first order of the asymptotic expansion, Chen (1997) showed that the application of the empirical likelihood density estimator, say $\tilde{f}_n$, in comparison to the simple kernel estimator f_n only changes the asymptotic mean squared error in second order. However, in second order the variance of $\tilde{f}_n$ is always smaller than the variance of f_n and the bias of the same asymptotic order is unchanged.

Now assume an independent and identically distributed sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of bivariate random variables has been observed, which is modelled by a homoscedastic nonparametric regression model

$$Y_i \;=\; m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.2}$$

where we are interested in the distribution F or density f of the centered i.i.d. errors $\varepsilon_1, \ldots, \varepsilon_n$. After estimating the conditional mean m in such a regression model one might be interested in the distribution of the observations around that mean, which is characterised by the distribution of the errors. Moreover an estimation of the error distribution function or density is the first step in analyzing features of that unknown distribution and is needed for hypotheses testing of certain qualities, such as symmetry, or testing the fit of a parametric class for the error distribution. For example, known symmetry or normality of the error distribution can lead to better efficiency of several statistical procedures in nonparametric regression [see Dette et al. (2002) and Neumeyer et al. (2006) for further references]. Furthermore estimated quantiles of the error distribution are needed to obtain prediction intervals for new observations [see Akritas and Van Keilegom (2001)]. Because the errors are unobserved, they have to be estimated by residuals $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$, $i = 1, \ldots, n$. Let, to this end, $\hat{m}$ denote some nonparametric regression function estimator. Denote by $\hat{F}_n$ the empirical distribution function for F and by $\hat{f}_n$ the kernel density estimator for f, both based on the residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$. See for instance Akritas and Van Keilegom (2001), Cheng (2004), Efromovich (2005), or Müller et al. (2007) for further motivation of error distribution estimation as well as asymptotic results on $\hat{F}_n$ and $\hat{f}_n$. Note that for $\hat{F}_n$ in comparison to F_n the asymptotic expansion changes due to the estimation of the regression function. The distribution estimator becomes biased and also the asymptotic variance changes in first order. The asymptotic variance of $\hat{F}_n$ can even be smaller than the asymptotic variance of F_n (however, F_n (and f_n) are not feasible in the model considered here).

Now assume again that we have additional information like before in terms of $E[g(\varepsilon_1)] = 0$, which should be incorporated into the estimation by the empirical likelihood method. For example, $g(y) = y$ yields the centeredness assumption, which is automatically valid due to the definition of the regression model, but so far is not incorporated explicitely into the estimation. Also a variance σ^2 might be known from previous experiments which leads to $g(y) = (y, y^2 - \sigma^2)^\top$. Analysis of empirical likelihood estimation in this regression context should be seen as starting point for further empirical likelihood procedures for testing for the hypothesis $E[g(\varepsilon_1)] = 0$ or even $E[g(\varepsilon_1|X_1) \mid X_1] = 0$ in heteroscedastic models. This will enable, for instance, to test for parametric structure of the conditional variance function by setting $g(\varepsilon|x) = \varepsilon^2 - \sigma^2_\vartheta(x)$, $\vartheta \in \Theta$.

Let $\overline{F}_n$ denote the empirical likelihood estimator for F based on the residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$ under additional information (1.1). For this estimator Kiwitt et al. (2008) showed that both asymptotic bias and variance of $\overline{F}_n$ are different in first order in comparison to the residual-based empirical distribution $\hat{F}_n$. The incorporation of the additional information can lead to an improved estimator; however, in contrast to the i.i.d.-case considered by Qin and Lawless (1994), it does not in all cases.

In the paper at hand we consider the empirical likelihood density estimator $\overline{f}_n$ based on the residuals and investigate whether in comparison to the residual-based kernel estimator $\hat{f}_n$ the asymptotic bias and variance change and whether an improvement of the estimation can be achieved. We show that in contrast to Chen (1997) the asymptotic mean squared error already changes in first order, due to a change in the bias, whereas the asymptotic variance only changes in second order. The paper is organized as follows. In section 2 we define the kernel density estimators as well as the empirical likelihood kernel density estimators for i.i.d. data and residuals, respectively. In section 3 we derive the asymptotic expansions for bias and variance of all four density estimators and compare the results. Section 4 discusses examples in theory as well as by means of a small simulation study. Technical assumptions are stated in an appendix.

2 Definition of the estimators

In the following we give a short motivation of kernel density estimation in order to have a starting point for motivation of the empirical likelihood density estimation below. Let $\varepsilon_1, \ldots, \varepsilon_n$ denote an absolutely continuous i.i.d. sample from density f, which is to be estimated. A first crude estimate of a density gives equal weight $1/n$ to each of n observations ε_i, $i = 1, \ldots, n$. To obtain a smooth estimator the kernel approach distributes the weight $1/n$ in the interval $[\varepsilon_i - h, \varepsilon_i + h]$ for some bandwidth h, such that each y obtains the weight $K((\varepsilon_i - y)/h)/h$ for some chosen density K with support $[-1, 1]$. Those weights are then added for $i = 1, \ldots, n$ to obtain the estimator

$$f_n(y) \;=\; \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{\varepsilon_i - y}{h}\Big).$$

Assumptions on the kernel K and bandwidth $h = h_n$ are postponed to the appendix for reason of better readability. Note that we apply a bandwidth with optimal rate $h = h_n \sim n^{-1/5}$ for kernel estimation of a twice continuously differentiable density.

We now assume that additional information about the underlying distribution is available. This auxiliary information is given in terms of equation (1.1), where $g = (g_1, \ldots, g_k)^\top : \mathbb{R} \to \mathbb{R}^k$ is a known function. Now instead of giving equal weight to all observations the empirical likelihood method [see Owen (2001)] gives weight $p_i \in [0, 1]$ to ε_i $(i = 1, \ldots, n)$. In order to include the information (1.1) into the estimation, the likelihood $\prod_{i=1}^n p_i$ (the probability that the given sample is observed) is maximized under the constraints $\sum_{i=1}^n p_i = 1$ (to obtain a probability distribution) and $\sum_{i=1}^n p_i g(\varepsilon_i) = 0$, which is the empirical version of (1.1). Now distributing such weights p_i in $[\varepsilon_i - h, \varepsilon_i + h]$ by a kernel approach as before gives the empirical likelihood kernel density estimator

$$\tilde{f}_n(y) \;=\; \frac{1}{h}\sum_{i=1}^{n} p_i K\Big(\frac{\varepsilon_i - y}{h}\Big)$$

as considered by Chen (1997).

Note that in our case the estimators f_n and $\tilde{f}_n$ are not available because $\varepsilon_1, \ldots, \varepsilon_n$ are not observable. We consider a nonparametric homoscedastic regression model (1.2) with independent observations, where the covariates $X_1, \ldots, X_n$ are i.i.d. with density f_X, and independent of the i.i.d. errors $\varepsilon_1, \ldots, \varepsilon_n$ with density f. For clarity reasons further technical model assumptions are listed in the appendix. In order to estimate the density f of the unobserved errors we build nonparametric residuals $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$, $i = 1, \ldots, n$, where $\hat{m}$ denotes the Nadaraya-Watson estimator [see Nadaraya (1964), Watson (1964)] for m, that is $\hat{m}(x) = (nb)^{-1}\sum_{i=1}^n k((X_i - x)/b)Y_i/\hat{f}_X(x)$ where $\hat{f}_X(x) = (nb)^{-1}\sum_{i=1}^n k((X_i - x)/b)$ is a kernel estimator for the covariate density. Assumptions on the kernel k and bandwidth b are again listed in the appendix. Note that we assume a bandwidth $b = b_n \sim n^{-1/5}$ of optimal rate for estimation of a twice continuously differentiable regression function. To estimate the error density we consider

$$\hat{f}_n(y) \;=\; \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{\hat{\varepsilon}_i - y}{h}\Big)$$

based on the residuals, instead of the non-feasible kernel estimator f_n.

Finally we define the empirical likelihood estimator

$$\overline{f}_n(y) \;=\; \frac{1}{h}\sum_{i=1}^{n} \hat{p}_i K\Big(\frac{\hat{\varepsilon}_i - y}{h}\Big)$$

with the same motivation as for $\tilde{f}_n$, but based on residuals. From Qin and Lawless (1994) and Kiwitt et al. (2008) it follows that the empirical likelihood weights $\hat{p}_i = 1/(n + n\hat{\eta}_n^\top g(\hat{\varepsilon}_i))$ where $\hat{\eta}_n$ is defined as solution of the equation $\sum_{i=1}^{n} g(\hat{\varepsilon}_i)/(1 + \hat{\eta}_n^\top g(\hat{\varepsilon}_i)) = 0$ such that $1 + \hat{\eta}_n^\top g(\hat{\varepsilon}_i) > n^{-1}$ for all $i = 1, \ldots, n$.

To compare the asymptotic performance of the estimators in the next section we give results for the asymptotic bias and variance of the four estimators f_n, $\tilde{f}_n$, $\hat{f}_n$, and $\overline{f}_n$.

3 Asymptotic results

3.1 (The kernel density estimator based on i.i.d. data.) From standard results in the kernel estimation literature [see Wand and Jones (1995), for instance] we have

$$\begin{aligned} E[f_n(y)] \;&=\; \frac{1}{h}\int K\Big(\frac{z-y}{h}\Big) f(z)\, dz \;=\; \int K(u) f(y+hu)\, du \\ &=\; f(y) + h^2 \frac{f''(y)}{2}\int u^2 K(u)\, du + o(h^2) \end{aligned}$$

by Taylor's expansion. For the variance one obtains similarly

$$\begin{aligned} \mathrm{Var}(f_n(y)) \;&=\; \frac{1}{nh^2}\mathrm{Var}\Big(K\Big(\frac{\varepsilon_1 - y}{h}\Big)\Big) \\ &=\; \frac{1}{nh^2}\int K^2\Big(\frac{z-y}{h}\Big) f(z)\, dz - \frac{1}{nh^2}\Big(\int K\Big(\frac{z-y}{h}\Big) f(z)\, dz\Big)^2 \\ &=\; \frac{1}{nh}\Big(f(y)\int K^2(u)\, du + O(h^2)\Big) - \frac{1}{n}\Big(f(y) + O(h^2)\Big)^2 \\ &=\; \frac{1}{nh} f(y)\int K^2(u)\, du - \frac{1}{n} f^2(y) + o\big(\frac{1}{n}\big). \end{aligned}$$

Thus the first order bias is of rate h^2, the first and second order variance terms are of rates $(nh)^{-1} \sim h^4$ and n^{-1}, respectively.

3.2 (The empirical likelihood kernel density estimator based on i.i.d. data) Chen (1997) has shown that

$$\begin{aligned} E[\tilde{f}_n(y)] &= E[f_n(y)] + o(\frac{1}{n}) \\ \mathrm{Var}(\tilde{f}_n(y)) &= \mathrm{Var}(f_n(y)) - \frac{1}{n} f^2(y) g(y)^\top \Sigma^{-1} g(y) + o(\frac{1}{n}). \end{aligned}$$

From the incorporation of the additional information by the empirical likelihood method there is no change of the bias, and no change of the variance in first order. However, the variance term of second order n^{-1} yields a reduction in comparison to f_n because by assumption $\Sigma = E[g(\varepsilon_1)g(\varepsilon_1)^\top]$ is positive definite (see appendix), and hence $g(y)^\top \Sigma^{-1} g(y) > 0$ (for all y).

3.3 (The residual-based kernel density estimator) In this subsection we will sketch the proof for a stochastic expansion of $\hat{f}_n$. In the following by $R_{n,\ell}$, $\ell = 1, \ldots, 4$, we denote remainder terms, each defined by the corresponding equality. Explanations for the expansions and for negligibility of the remainder terms are given below. We obtain for the residual-based kernel density estimator that

$$\begin{aligned} \hat{f}_n(y) &= \frac{1}{nh} \sum_{i=1}^n K\Big(\frac{\varepsilon_i - y}{h}\Big) + \frac{1}{nh^2} \sum_{i=1}^n K'\Big(\frac{\varepsilon_i - y}{h}\Big)(\varepsilon_i - \hat{\varepsilon}_i) + R_{n,1} && (3.1) \\ &= f_n(y) + \frac{1}{nh^2} \sum_{i=1}^n K'\Big(\frac{\varepsilon_i - y}{h}\Big) \frac{1}{nb} \sum_{j=1}^n k\Big(\frac{X_i - X_j}{b}\Big) \frac{Y_j - m(X_i)}{f_X(X_i)} + R_{n,2} && (3.2) \\ &= f_n(y) + \frac{1}{n} \sum_{j=1}^n \varepsilon_j \int \frac{1}{h^2} K'\Big(\frac{z-y}{h}\Big) f(z)\, dz \int \frac{1}{b} k\Big(\frac{u - X_j}{b}\Big)\, du + B_n + R_{n,3} && (3.3) \\ &= f_n(y) + f'(y) \frac{1}{n} \sum_{j=1}^n \varepsilon_j + B_n + R_{n,4} && (3.4) \end{aligned}$$

with the deterministic term

$$B_n = \int \frac{1}{h^2} K'\Big(\frac{z-y}{h}\Big) f(z)\, dz \int \frac{1}{b} k\Big(\frac{u-v}{b}\Big)(m(v) - m(u)) f_X(v)\, dv = b^2 f'(y) B + o(b^2).$$

where $B = \frac{1}{2} \int (m f_X)''(x) - (m f_X'')(x)\, dx \int u^2 k(u)\, du$. To obtain the expansion for B_n we have used calculations typical for kernel estimation theory and also that $\int h^{-2} K'((z-y)/h) f(z)\, dz = \int h^{-1} K((z-y)/h) f'(z)\, dz = f'(y) + O(h^2)$ by integration by parts and Taylor's expansion. One can show that under the assumptions stated in the appendix the variance of the remainder term $R_{n,4}$ in (3.4) is of order $o(1/n)$ and the covariance of $R_{n,4}$ with the other terms in the expansion (3.4) of $\hat{f}_n(y)$ is (by an application of Cauchy-Schwarz' inequality) also of order $o(1/n)$. To this end note that $R_{n,4} = R_{n,1} + (R_{n,2} - R_{n,1}) + (R_{n,3} - R_{n,2}) + (R_{n,4} - R_{n,3})$. Here with a Taylor expansion of the kernel K up to order five we obtain negligibility of $R_{n,1}$, whereas the term $R_{n,2} - R_{n,1}$ is discussed by inserting the definition of $\hat{m}$ in $\varepsilon_i - \hat{\varepsilon}_i = \hat{m}(X_i) - m(X_i)$ in (3.1) (see section 2) and by replacing the random denominator $\hat{f}_X$ by the true density f_X. Simple but cumbersome calculations of variances yield negligibility of $R_{n,3} - R_{n,2}$ [inserting the model definition $Y_j = \varepsilon_j - m(X_j)$ in (3.2)] and $R_{n,4} - R_{n,3}$ [by evaluation of the integrals in (3.3)]. Technical details are omitted for the sake of brevity. From (3.4) and the results in section 3.1 we have

$$E[\hat{f}_n(y)] = f(y) + h^2 \frac{f''(y)}{2} \int u^2 K(u)\, du + b^2 f'(y) B + o(h^2) + o(b^2) \tag{3.5}$$

$$\begin{aligned} \operatorname{Var}(\hat{f}_n(y)) &= \operatorname{Var}\Big(f_n(y) + f'(y) \frac{1}{n} \sum_{j=1}^n \varepsilon_j \Big) + o(\frac{1}{n}) \\ &= \operatorname{Var}(f_n(y)) + \frac{1}{n} \operatorname{Var}(\varepsilon_1)(f'(y))^2 + \frac{2}{n} f'(y) \int \frac{1}{h} K\Big(\frac{z-y}{h}\Big) z f(z)\, dz + o(\frac{1}{n}) \\ &= \frac{1}{nh} f(y) \int K^2(u)\, du + \frac{1}{n}\Big(\operatorname{Var}(\varepsilon_1)(f'(y))^2 + 2y f(y) f'(y) - f^2(y) \Big) + o(\frac{1}{n}). \end{aligned}$$

In comparison to the asymptotic expectation and variance of $f_n(y)$ as stated in section 3.1 we see that the estimation of the regression function m results in a change of the bias of (first) order b^2 and the variance of (second) order n^{-1}.

3.4 (The residual-based empirical likelihood kernel density estimator) Finally, we consider the residual-based empirical likelihood density estimator, for which by the definition of the weights $\hat{p}_i = n^{-1}(1 - \hat{\eta}_n^\top g(\hat{\varepsilon}_i) + (\hat{\eta}_n^\top g(\hat{\varepsilon}_i))^2 / (1 + \hat{\eta}_n^\top g(\hat{\varepsilon}_i)))$ we have

$$\overline{f}_n(y) = \hat{f}_n(y) - \hat{\eta}_n^\top \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i) K\Big(\frac{\hat{\varepsilon}_i - y}{h}\Big) + R_{n,5} \tag{3.6}$$

for a remainder term $R_{n,5}$, which is of order $O_p((nh)^{-1})$ because $\hat{\eta}_n = O_p(n^{-1/2})$; see Lemma B.4(i) in Kiwitt et al. (2008). From Proposition 3.2 and Lemma B.2(ii) in the same reference we have

$$\hat{\eta}_n = \Sigma^{-1} \left(\frac{1}{n} \sum_{i=1}^n (g(\varepsilon_i) - \varepsilon_i E[g'(\varepsilon_1)]) - b^2 E[g'(\varepsilon_1)] B \right) + o(b^2) + o_p(\frac{1}{\sqrt{n}}) \tag{3.7}$$

where $\Sigma = E[g(\varepsilon_1) g(\varepsilon_1)^\top]$ and B is defined in section 3.3. (see also the remark below the list of assumptions in the appendix). Using Taylor's expansion in a similar way as in section 3.3 one can show that

$$\frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i) K\Big(\frac{\hat{\varepsilon}_i - y}{h}\Big) = \frac{1}{nh} \sum_{i=1}^n g(\varepsilon_i) K\Big(\frac{\varepsilon_i - y}{h}\Big) + R_{n,6} \tag{3.8}$$

with some negligible remainder term $R_{n,6}$. Further,

$$\frac{1}{nh} \sum_{i=1}^n g(\varepsilon_i) K\Big(\frac{\varepsilon_i - y}{h}\Big) = \int g(y) K\Big(\frac{z-y}{h}\Big) f(z)\, dz + O_p(\frac{1}{\sqrt{nh}})$$

$$= g(y)f(y) + O(h^2) + O_p(\frac{1}{\sqrt{nh}}), \tag{3.9}$$

and from (3.6) together with (3.7) and (3.8), (3.9) we obtain an expansion

$$\begin{aligned} \overline{f}_n(y) &= \hat{f}_n(y) - f(y)\frac{1}{n}\sum_{i=1}^{n}(g(\varepsilon_i)^\top - \varepsilon_i E[g'(\varepsilon_1)^\top])\Sigma^{-1}g(y) \\ &\quad + b^2 f(y)BE[g'(\varepsilon_1)^\top]\Sigma^{-1}g(y) + o(b^2) + R_{n,7} \end{aligned} \tag{3.10}$$

for some remainder term $R_{n,7}$. One can show with lengthy but simple calculations that the variance of the remainder term is of order $o(1/n)$ and the covariance of $R_{n,7}$ with all other terms in the expansion is (by an application of Cauchy-Schwarz' inequality) also of order $o(1/n)$. Hence, we obtain from (3.10) and (3.5) that

$$\begin{aligned} E[\overline{f}_n(y)] &= E[\hat{f}_n(y)] + b^2 f(y)BE[g'(\varepsilon_1)^\top]\Sigma^{-1}g(y) + o(b^2) \\ &= f(y) + h^2\frac{f''(y)}{2}\int u^2 K(u)\,du + b^2 B\Big(f'(y) + f(y)E[g'(\varepsilon_1)^\top]\Sigma^{-1}g(y)\Big) \\ &\quad + o(h^2) + o(b^2) \end{aligned} \tag{3.11}$$

and from (3.10) and (3.4) that

$$\begin{aligned} \mathrm{Var}(\overline{f}_n(y)) &= \mathrm{Var}(\hat{f}_n(y)) \\ &\quad + \frac{1}{n}f^2(y)E\Big[\Big(g(\varepsilon_2)^\top - E[g'(\varepsilon_1)^\top]\varepsilon_2\Big)\Sigma^{-1}g(y)\Big(g(\varepsilon_2)^\top - E[g'(\varepsilon_1)^\top]\varepsilon_2\Big)\Big]\Sigma^{-1}g(y) \\ &\quad - \frac{2}{n}f^2(y)\Big(g(y)^\top - E[g'(\varepsilon_1)^\top]y\Big)\Sigma^{-1}g(y) \\ &\quad - \frac{2}{n}f(y)f'(y)E\Big[\Big(g(\varepsilon_2)^\top - E[g'(\varepsilon_1)^\top]\varepsilon_2\Big)\varepsilon_2\Big]\Sigma^{-1}g(y) + o(\frac{1}{n}). \end{aligned} \tag{3.12}$$

In comparison to expectation and variance of $\hat{f}_n(y)$ as given in section 3.3 we see that in contrast to Chen's (1997) results the empirical likelihood method for the residual-based estimators leads to a change in the bias in first order. This change however only arises in the bias of order b^2, that is due to the estimation of the residuals, whereas the h^2-bias remains as before. In the variance we observe a change only in second order. We will investigate in section 4 if this change means that the estimation can be improved by the application of the empirical likelihood method.

3.5 (The mean squared errors) To compare the asymptotic mean squared errors in second order we need more assumptions to obtain the second order bias. For simplicity in addition to the assumptions stated in the appendix we assume f, m and f_X to be thrice continuously differentiable. Because the kernels K and k are symmetric with compact supports the third moments of the kernels vanish, and hence the bias remainder terms $o(h^2)$ and $o(b^2)$ are of order $o(h^3)$ and $o(b^3)$, respectively. Then, we obtain for the mean squared errors (note that $h^5 \sim n^{-1}$, $b^5 \sim n^{-1}$ and $(nh)^{-1} \sim h^4 \sim n^{-4/5}$, $(nb)^{-1} \sim b^4 \sim n^{-4/5}$) that

$$\begin{aligned}
\text{mse}(f_n(y)) &= \frac{1}{nh} f(y) \int K^2(u)\,du + h^4 \Big(\frac{f''(y)}{2} \int u^2 K(u)\,du\Big)^2 - \frac{1}{n} f^2(y) + o(\frac{1}{n}) \\
\text{mse}(\tilde{f}_n(y)) &= \text{mse}(f_n(y)) - \frac{1}{n} f^2(y) g(y)^\top \Sigma^{-1} g(y) + o(\frac{1}{n}) \\
\text{mse}(\hat{f}_n(y)) &= \text{mse}(f_n(y)) + b^2 h^2 f''(y) f'(y) B \int u^2 K(u)\,du \\
&\quad + \frac{1}{n}\Big(\text{Var}(\varepsilon_1)(f'(y))^2 + 2y f(y) f'(y)\Big) + b^4 (f'(y))^2 B^2 + o(\frac{1}{n}) \\
\text{mse}(\overline{f}_n(y)) &= \text{mse}(\hat{f}_n(y)) + \Big(b^2 f(y) B E[g'(\varepsilon_1)^\top] \Sigma^{-1} g(y)\Big)^2 \\
&\quad + b^2 f(y) B E[g'(\varepsilon_1)^\top] \Sigma^{-1} g(y) \Big(h^2 f''(y) \int u^2 K(u)\,du + 2b^2 B f'(y)\Big) \\
&\quad + \frac{1}{n} f^2(y) E\Big[\Big(g(\varepsilon_2)^\top - E[g'(\varepsilon_1)^\top]\varepsilon_2\Big) \Sigma^{-1} g(y) \Big(g(\varepsilon_2)^\top - E[g'(\varepsilon_1)^\top]\varepsilon_2\Big)\Big] \Sigma^{-1} g(y) \\
&\quad - \frac{2}{n} f^2(y) \Big(g(y)^\top - E[g'(\varepsilon_1)^\top] y\Big) \Sigma^{-1} g(y) \\
&\quad - \frac{1}{n} f(y) f'(y) E\Big[\Big(g(\varepsilon_2)^\top - E[g'(\varepsilon_1)^\top]\varepsilon_2\Big)\varepsilon_2\Big] \Sigma^{-1} g(y) + o(\frac{1}{n}).
\end{aligned}$$

4 Examples and simulations

It was shown before that the residual-based kernel density estimator and the residual-based empirical likelihood kernel density estimator differ in the asymptotic bias in first and in the asymptotic variance in second order. However, the application of the empirical likelihood weights does not necessarily lead to an improvement. Therefore, the theoretical asymptotic results are considered for two typical examples of having additional information.

Afterwards the behaviour of both estimators for smaller sample sizes is observed in a small simulation study.

4.1 (Additional information: centered errors) The model assumption of centered errors is to be explicitly included in the estimation. By choosing $g(\varepsilon) = \varepsilon$ the variance formula (3.12) reduces with $E[g'(\varepsilon)] = 1$, due to $g(y) - E[g'(\varepsilon_1)^\top] y = y - 1 \cdot y = 0$, to $\text{Var}(\overline{f}_n(y)) = \text{Var}(\hat{f}_n(y)) + o(n^{-1})$. Hence, there is no second order improvement in the variance possible, but for the asymptotic bias follows by taking $\Sigma = E[\varepsilon^2] = \sigma^2$ into account [see (3.11) under the assumptions of section 3.5]

$$E[\overline{f}_n(y)] = f(y) + h^2 \frac{f''(y)}{2} \int u^2 K(u)\,du + b^2 B\Big(f'(y) + f(y)\frac{1}{\sigma^2} y\Big) + o(h^3) + o(b^3).$$

Thus a first order improvement in the asymptotic bias is possible. In case of normally distributed errors follows with $f'(y) = -y f(y)/\sigma^2$ that the b^2-bias due to the estimation of the regression function cancels completely. Figure 1 illustrates an improvement for most y, but nevertheless, for some points the residual-based kernel density estimator has the smaller bias. Here only the dominating terms of first and second order are depicted and we take $h = c_1 n^{-1/5}$, $b = c_2 n^{-1/5}$, where, for simplicity, $c_1^2 \int u^2 K(u)\,du = 1$ and $c_2^2 B = 1$. Integrating the differences in the bias about all y gives an overall improvement in the asymptotic bias, and hence in the asymptotic mean integrated squared error (amise). To see this note that from the formulae in section 3.5 we obtain by symmetry of f and because $f'(y) = -yf(y)/\sigma^2$ that

$$\begin{aligned}\text{amise}(\overline{f}_n) &= \text{amise}(\hat{f}_n) + \frac{b^2h^2B}{\sigma^2}\int yf(y)f''(y)\,dy \int u^2K(u)\,du \\ &\quad + \frac{2b^4B^2}{\sigma^2}\int yf(y)f'(y)\,dy + \frac{b^4B^2}{\sigma^4}\int y^2 f^2(y)\,dy \\ &= \text{amise}(\hat{f}_n) + \frac{b^4B}{\sigma^2}\int yf(y)f'(y)\,dy \;=\; \text{amise}(\hat{f}_n) - \frac{b^4B^2}{4\sqrt{\pi}\sigma^3}.\end{aligned}$$

The greatest improvement is obtained in cases of a large bias $b^2|B|$ due to regression estimation combined with small variances σ^2.

FIGURE 1 HERE

4.2 (Additional information: centered errors and Var$(\varepsilon) = \sigma^2$) In some application the variance of the errors σ^2 is known. This information and the centeredness of the errors can be included in the estimation by defining $g(\varepsilon)^\top = (\varepsilon,\ \varepsilon^2 - \sigma^2)$. For convenience we assume that the third moments of the error distribution are zero. With $E[g'(\varepsilon)] = (1, 0)^\top$ and $\Sigma = \text{diag}(\sigma^2, E[(\varepsilon^2-\sigma^2)^2])$ follows the same asymptotic change in the bias as in section 3.1, whereas (3.12) yields

$$\text{Var}(\overline{f}_n(y)) \;=\; \text{Var}(\hat{f}_n(y)) - \frac{1}{n}f^2(y)\frac{1}{E[(\varepsilon^2-\sigma^2)^2]}(y^2-\sigma^2)^2 + o\left(n^{-1}\right).$$

Because of the obviously negative additional term, there is always a second order improvement for all y in the asymptotic variance of the residual-based empirical likelihood kernel density estimator. In Figure 2 you see the improvements in case of standard normally distributed errors. In that case for the asymptotic mean integrated squared error one obtains

$$\text{amise}(\overline{f}_n) \;=\; \text{amise}(\hat{f}_n) - \frac{b^4B^2}{4\sqrt{\pi}\sigma^3} - \frac{3}{16\sqrt{\pi}\sigma n}.$$

FIGURE 2 HERE

4.3 (Simulation study) In the simulation study the mean squared error of both estimators is analyzed for sample sizes $n = 50, 100, 250$. To this end the mean squared error of the residual-based empirical likelihood density estimator is approximated by $\hat{\text{mse}}(\overline{f}_n(y)) = \frac{1}{N}\sum_{i=1}^{N}\left(\overline{f}_{n,i}(y) - f(y)\right)^2$, where $N = 10000$ is the number of repetitions for data generation and $\overline{f}_{n,i}(y)$ is the estimator applied to the i-th sample. $\hat{\text{mse}}(\hat{f}_n(y))$ is defined analogously. The regression function $m(x) = 5x^2$ is assumed and estimated by the Nadaraya-Watson estimator with bandwidth $b = n^{-\frac{1}{5}}$. For the kernel density estimator we choose the bandwidth $h = n^{-\frac{1}{5}}$, and for both estimators the Epanechnikov kernel. Figure 3 shows the simulations for standard normally distributed errors and the two-dimensional additional information $g(\varepsilon)^\top = (\varepsilon,\ \varepsilon^2 - \sigma^2)$. For all sample sizes mentioned above the residual-based kernel density estimator is improved for almost all y by the empirical likelihood weights. The greatest improvements are in the interval $[-0.2, 1]$, which is supported by the theoretical results [see panel (3) in Figure 2]. In Figure 4 are some other examples presented, where the main consequence is that the residual-based empirical likelihood kernel density estimator has in all cases, for at least the most y, the smaller mean squared error.

FIGURES 3 AND 4 HERE

A Assumptions

1. The univariate covariates $X_1, \dots, X_n$ are independent and identically distributed with distribution function F_X on compact support, say $[0, 1]$. F_X has an in $(0, 1)$ twice continuously differentiable density f_X, such that $\inf_{x\in[0,1]} f_X(x) > 0$. The regression function m is twice continuously differentiable in $(0, 1)$ with bounded derivatives.

2. The errors $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed with distribution function F. They are centered, $E[\varepsilon_1] = 0$, with variance $\sigma^2 = \text{Var}(\varepsilon_1) \in (0, \infty)$, existing third moment, and are independent from the covariates. F is continuously differentiable with bounded, everywhere positive density f. There exist constants $\gamma, C > 0$ and $\beta \in (0, 4)$ such that $|F(y + z) - F(y) - zf(y)| \le C|z|^{1+\beta}$ for all $z \in \mathbb{R}$ with $|z| \le \gamma$.

3. Let k denote a twice continuously differentiable symmetric density with compact support and $\int uk(u)\,du = 0$. Let $b = b_n$ be a sequence of bandwidths such that $b \sim n^{-1/5}$.

4. We assume that g_j is continuously differentiable with $E[g_j^2(\varepsilon_1)] < \infty$ and $E[|g_j'(\varepsilon_1)|] < \infty$, and there exist constants γ, C and $\beta > 0$ such that $|\int (g_j(y + z) - g_j(y) - zg_j'(y))f(y)\,dy| \le C|z|^{1+\beta}$ for all $z \in \mathbb{R}$ with $|z| \le \gamma$, $j = 1, \dots, k$. We assume that $\min_{1\le i\le n} g_j(\hat{\varepsilon}_i) < 0 < \max_{1\le i\le n} g_j(\hat{\varepsilon}_i)$ for all $j = 1, \dots, k$, and that $\Sigma = E[g(\varepsilon_1)g(\varepsilon_1)^\top]$ and $\sum_{i=1}^{n} g(\hat{\varepsilon}_i)g(\hat{\varepsilon}_i)^\top$ are

positive definite. We assume the existence of constants δ, C such that for some positive $\kappa < 4$ and all $j = 1, \ldots, k$, $(E[\sup_{\substack{z,\tilde{z}\in\mathbb{R}:|z|\leq\delta,\\ |\tilde{z}|\leq\delta,|z-\tilde{z}|\leq\xi}} (g_j(\varepsilon_1 + z) - g_j(\varepsilon_1 + \tilde{z}))^2])^{1/2} \leq C\xi^{1/\kappa}$.

5. We assume that $\sup_{x\in\mathbb{R}} |(g_j f)(x)| < \infty$, $\sup_{x\in\mathbb{R}} |(g_j f)'(x)| < \infty$, $\sup_{x\in\mathbb{R}} |(g_j^2 f)(x)| < \infty$ and $\sup_{x\in\mathbb{R}} |(g_j^2 f)'(x)| < \infty$ for all $j = 1, \ldots, k$.

6. Let K denote a five times continuously differentiable symmetric density with support $[-1, 1]$, $K(-1) = K(1) = 0$ and $\int uK(u)\,du = 0$. Let $h = h_n$ be a sequence of bandwidths such that $h \sim n^{-1/5}$.

See Kiwitt et al. (2008) for interpretations of assumptions 1–4 concerning the regression model, estimation of the regression function and the empirical likelihood procedure. Note that the assumption $nb^4 = O(1)$ by Kiwitt et al. (2008) is not valid under our assumptions, but an inspection of the proof of Proposition 3.2 and Lemma B.2(ii) in that reference shows that (3.7) can also be obtained under $nb^5 = O(1)$. Moreover, Kiwitt et al.'s (2008) bandwidth assumptions $nb^{3+2\alpha}(\log(h^{-1}))^{-1} \to \infty$ (for some $\alpha > 0$) and $n^\beta b^{1+\beta}(\log(h^{-1}))^{-1-\beta} \to \infty$ with β as in assumption 4 are valid for our choice of bandwidth rate (for every $\alpha \in (0, 1)$).

The additional technical assumption 5 is needed to bound some remainder terms.

Assumption 6 is due to the kernel density estimation. The high rate of smoothness needed for the kernel is due to bounding of remainder terms in a Taylor expansion with the aim of obtaining second order rates.

B References

M. Akritas, I. Van Keilegom (2001). *Nonparametric estimation of the residual distribution.* Scand. J. Statist. 28, 549–567.

S. X. Chen (1997). *Empirical-likelihood-based kernel density estimation.* Austral. J. Statist. 39, 47–56.

F. Cheng (2004). *Weak and strong uniform consistency of a kernel error density estimator in nonparametric regression.* J. Statist. Plann. Inference 119, 95–107.

H. Dette, S. Kusi-Appiah, N. Neumeyer (2002). *Testing symmetry in nonparametric regression models* Nonparam. Statist. 14, 477–494.

T. DiCiccio, P. Hall, J. P. Romano (1989). *Comparison of parametric and empirical likelihood functions.* Biometrika 76, 465–476.

S. Efromovich (2005). *Estimation of the density of regression errors.* Ann. Statist. 33, 2194–2227.

J. H. J. Einmahl, I. W. McKeague (2003). *Empirical likelihood based hypothesis testing.* Bernoulli 9, 267–290.

P. Hall, B. LaScala (1990). *Methodology and algorithms of empirical likelihood.* Internat. Statist. Rev. 58, 109–127.

N. L. Hjort, I. W. McKeague, I. Van Keilegom (2009). *Extending the scope of empirical likelihood.* Ann. Statist. 37, 1079–1111.

Y. Kitamura (1997). *Empirical Likelihood Methods with weakly dependent processes.* Ann. Statist. 25, 2084–2102.

S. Kiwitt, E.-R. Nagel, N. Neumeyer (2008). *Empirical likelihood estimators for the error distribution in nonparametric regression models.* Mathem. Meth. Statist. 17, 241–260.

U. Müller, A. Schick, W. Wefelmeyer (2007). *Estimating the error distribution function in semiparametric regression.* Statist. Decisions 25, 1–18.

É. A. Nadaraya (1964). *On nonparametric estimates of density functions and regression curves.* J. Probab. Appl. 10, 186–190.

N. Neumeyer, H. Dette and E.–R. Nagel (2006). *Bootstrap tests for the error distribution in linear and nonparametric regression models.* Austr. N. Zeal. J. Statist. 48, 129–156.

A. B. Owen (2001). *Empirical Likelihood.* Chapman & Hall.

J. Qin, J. Lawless (1994). *Empirical likelihood and general estimating equations.* Ann. Statist. 22, 300–325.

M. P. Wand, M. C. Jones (1995). *Kernel smoothing.* Chapman & Hall.

G. S. Watson (1964). *Smooth Regression Analysis.* Sankhya A 26, 359–372.

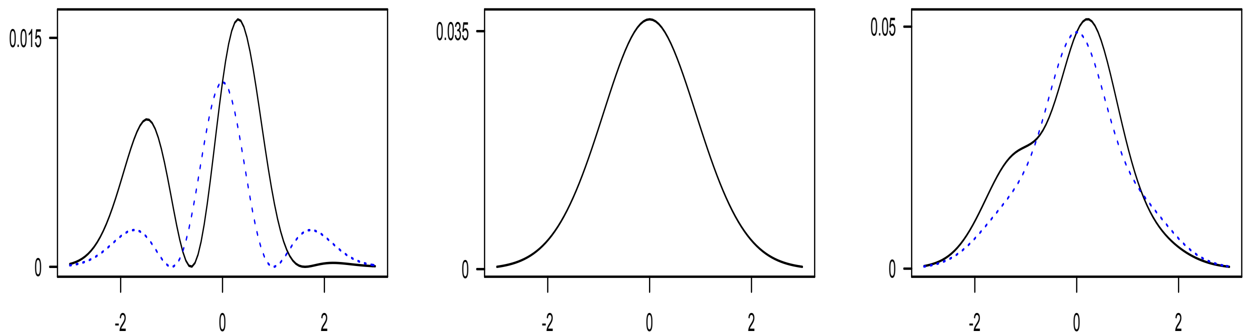

Figure 1: The figure shows the asymptotic bias (panel 1), the asymptotic variance (panel 2) and the asymptotic mean squared error (panel 3) of the residual-based kernel density estimator (solid line) and the residual-based empirical likelihood kernel density estimator (dashed line) in case of standard normally distributed errors and $g(\varepsilon) = \varepsilon$ with $n = 25$.

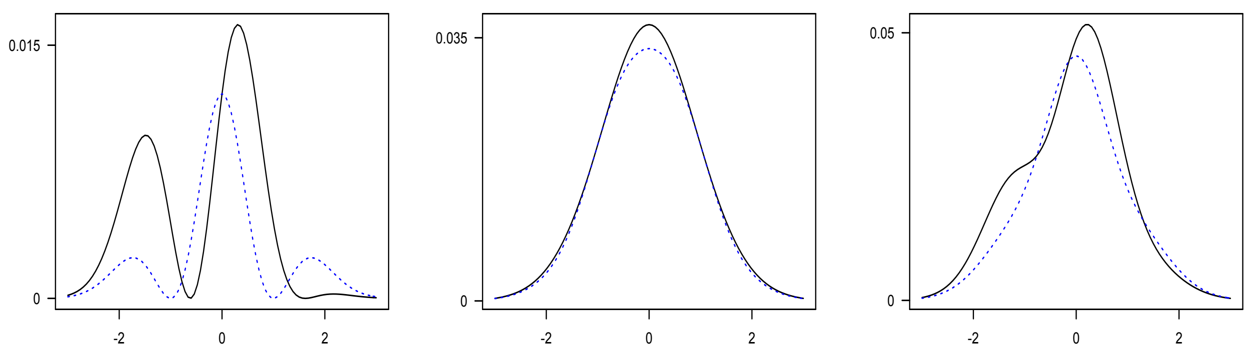

Figure 2: The figure shows the asymptotic bias (panel 1), the asymptotic variance (panel 2) and the asymptotic mean squared error (panel 3) of the residual-based kernel density estimator (solid line) and the residual-based empirical likelihood kernel density estimator (dashed line) in case of standard normally distributed errors and $g(\varepsilon)^\top = (\varepsilon, \varepsilon^2 - \sigma^2)$ with $n = 25$.

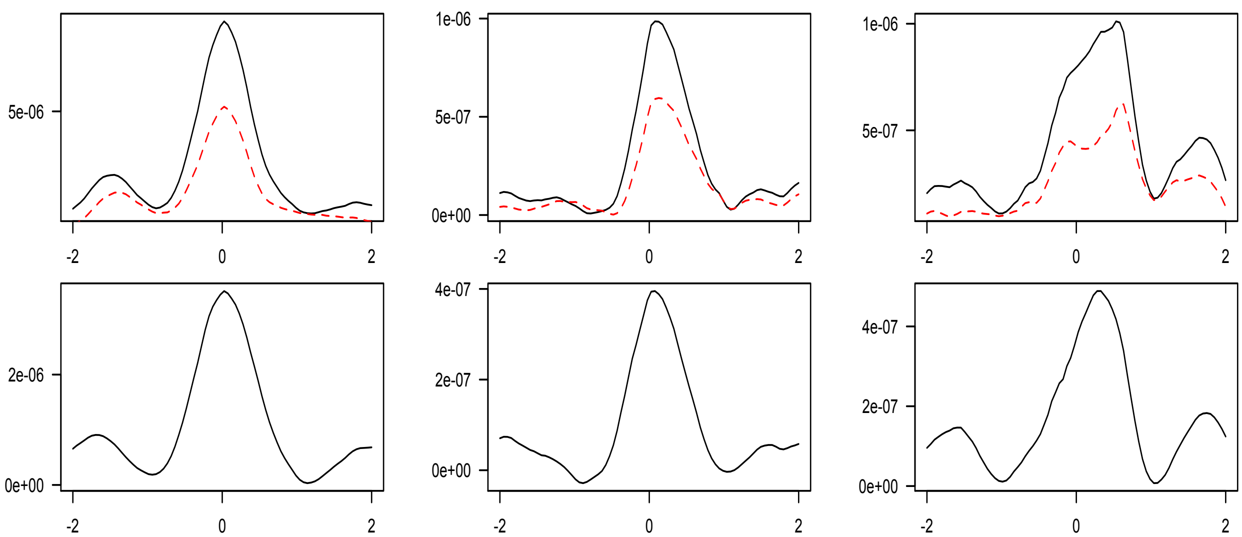

Figure 3: The upper row shows the approximated mean squared error of $\overline{f}_n(y)$ (dashed line) and $\hat{f}_n(y)$ (solid line) for $g(\varepsilon)^\top = (\varepsilon, \varepsilon^2 - \sigma^2)$ with $n = 50,\ 100,\ 250$ and standard normally distributed errors. The lower row shows the difference $\hat{\text{mse}}(\hat{f}_n(y)) - \hat{\text{mse}}(\overline{f}_n(y))$ of the two upper curves.

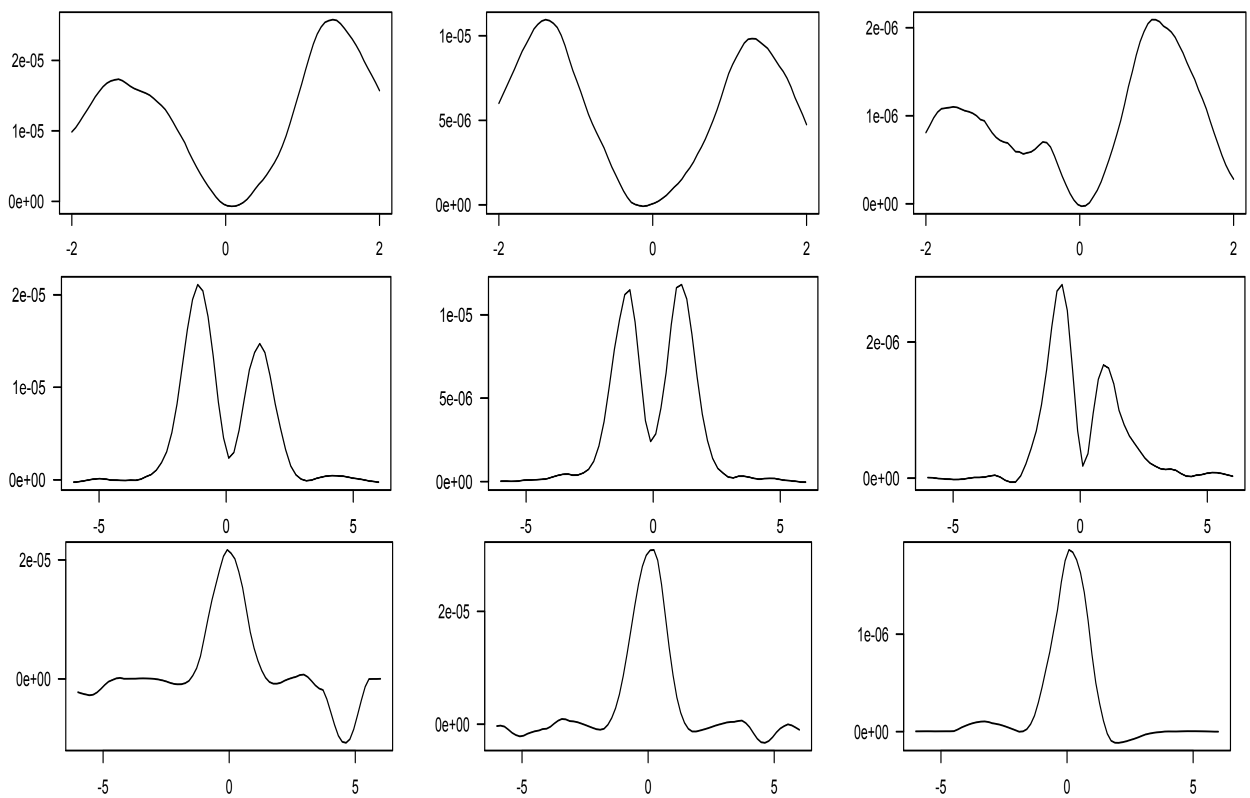

Figure 4: All rows show the difference $\hat{\text{mse}}(\hat{f}_n(y)) - \hat{\text{mse}}(\overline{f}_n(y))$ for $n = 50,\ 100,\ 250$. The first one illustrates the case of standard normally distributed errors with $g(\varepsilon) = \varepsilon$. The other two consider the case of Student's t-distributed errors with 3 degrees of freedom and respectively $g(\varepsilon) = \varepsilon$, $g(\varepsilon)^\top = (\varepsilon,\ \varepsilon^2 - \sigma^2)$.